Using Survey Data to Estimate the Population Size and Distribution of MSM

Joseph A. Catania, Jesse Canchola, Lance Pollack

November 2002

Joseph A. Catania is a Professor in the Department of Medicine at the University of California, San Francisco. Jesse Canchola is an Analyst at UCSF. Lance Pollack is a Research Specialist at UCSF. They are with the Health Survey Research Unit/Center for AIDS Prevention Studies/Department of Medicine, University of California, San Francisco.

Address correspondence to Joseph A. Catania, PhD, Center for AIDS Prevention Studies, University of California, San Francisco, 74 New Montgomery, Suite 600, San Francisco, CA 94105, jcatania@psg.ucsf.edu.

This work was supported by Grants MH51523, MH42459, MH43892, and MH46240 from NIMH and NIA to the first author. We would like to thank Juan Ruiz, MD, MPH, and Arthur Johnson, BA, currently with the California State Office of AIDS, and Richard K. P. Sun, MD, MPH, formerly with the California State Office of AIDS, for ideas and State AIDS data.

Abstract

We compared two methods of estimating the size of the adult men who have sex with men (MSM) population: "probability survey" and "never married" methods. The two methods produced similar overall population size estimates for our test case, California, and were distributed proportionately across geographic areas in relation to number of AIDS cases among MSM. The "probability survey" method, relative to the "never married" method, produced consistently, but only modestly larger population size estimates across geographic areas (except rural regions). However, an advantage of the "probability survey" method is that it can be used to gather additional information useful for estimating the size of the adolescent MSM population. Also, the 95% confidence intervals may be used to adjust for under-reporting and for survey budgeting purposes. We discuss problems related to self-disclosure of sexual orientation in surveys, and present data suggesting that this problem may not be as substantial as previously thought. In summary, the survey method is recommended because of its greater flexibility.

Introduction

To budget for and conduct probability-based surveys of men who have sex with men (MSM), it is important to have reasonable population size and distribution estimates for a given geographic area. In addition, to estimate the magnitude of mental health, substance abuse, and physical health problems among MSM, an accurate estimation of the MSM population size provides the needed denominator for guiding public health efforts. Because the U.S. Census does not ask questions about sexual orientation, we must rely more heavily on estimating techniques from which we infer the size and distribution of the MSM population.

Estimates of the MSM population based on probability sample surveys are subject to nonresponse distortion and are usually considered to be underestimates of the true population (Binson et al. 1995). Probability surveys also are costly and are, therefore, conducted infrequently at state or local levels. The strength of the probability survey method is that sexual orientation is determined by asking people directly, rather than through inference. Furthermore, statistical confidence intervals may be derived for point estimates, which can be used to adjust MSM population size estimates. In addition, extrapolation from survey data is based on a straightforward generalization of estimators of the MSM population.

An alternative, the "never married" method, makes inferences on marital status from census data by estimating the number of MSM based on the proportion of men over the age of 44 years that have never been married (Holmberg 1996). This method, however, requires various adjustments. For instance, since large numbers of young men that have never married are heterosexuals, the "never married" method generalizes the estimate of never being married from the 44 years and over category and applies the estimate to men under the age of 44. This extrapolation assumes that men in the older age group provide a better estimate of the proportion of adult MSM

unconfounded by large numbers of never married heterosexual men, as occurs among young adult populations. Also, additional adjustment is needed because African American men are more likely to never marry than White men. Further, this method is unable to estimate the MSM adolescent population. Nevertheless, the strength of the "never married" method is that it uses basic U.S. Census information, which may be less likely to be under-reported in the general population.

The present investigation compares the "never married" and the "probability survey" methods for estimating the proportionate distribution and population size for MSM. Using California as a case example, we compared the estimates obtained from these two methods with the known distribution of AIDS cases in adult MSM. The discussion section provides MSM population size estimates for California MSM adolescents, along with national estimates.

Methods

Probability survey method**.** Similar to Binson et al. (1995), and the recent Urban Men's Health Study (Blair 1999; Catania et al. 2001), we based our estimates of adult (18 years and older) MSM (per geographic areas) on weighted data aggregated across multiple waves (seven) of the General Social Survey (GSS) for the following years: 1988-1991, 1993-1994, and 1996. The GSS is a national probability survey(s) of the general adult population of the United States. MSM were defined by survey questions that ask for the gender of sexual partners since age 18, and only men who reported male partners were classified as MSM. The process of aggregating population estimates across methodologically compatible surveys increases the reliability of the point estimates. The GSS population estimates were then multiplied by the most recent population size estimates from the 1999 Current Population Survey (CPS) for California. This approach then generalizes national estimates to a large populous state. The use of the 1999 CPS

survey has advantages over the 1990 U.S. Census because it provides estimates of population size that are corrected for migration and other demographic changes in the population since 1990 (see table notes, Table 1). The prevalence of MSM was broken into six geographic categories in the United States: 1) the central city portion of the 12 largest metropolitan statistical areas (MSAs); 2) the central city portion of the next 88 largest MSAs; 3) the non-central city portion of the 12 largest MSAs; 4) the non-central city portion of the next 88 largest MSAs; 5) other urban (counties having towns of 10,000 or more population); and 6) other rural (counties having no towns of 10,000 or more population). California MSAs that are among the 12 largest in the United States (specifically the Los Angeles-Riverside-Orange County MSA and the San Francisco-Oakland-San Jose MSA) provided the population numbers for categories 1 and 3 (depending on whether location was marked as central city). Similarly, California MSAs that were among the next 88 largest MSAs in the United States (a total of nine MSAs) provide the population numbers for categories 2 and 4. We used Table 43 from the 1995 Statistical Abstract of the United States to make this determination. For category 5 we could not readily make the same distinction as did GSS, so we included in this category all Californians living in MSAs not among the 100 largest in the United States (five MSAs). Category 6 included all Californians not living in an MSA.

Never married method. Holmberg developed the most sophisticated application of the "never married" method for estimating the MSM population (Holmberg, 1996). Based on the 1999 CPS, we estimated the proportion of men over the age of 44 years who have never been married (by geographic area in California). Since African American men are more likely than White men to have never been married, the proportion of never married men over age 44 was adjusted for the proportion in each geographic area estimated to be African American. The proportion of never married men over the age of 44 was multiplied by the number of men aged 18 to 44 to

estimate the proportion of men in the younger age group who are MSM. Holmberg (1996) has previously reported excellent correspondence between the estimated proportions of MSM using the never married method and the proportion of MSM AIDS cases across geographic areas. The estimating formula and adjustments are further described in the table notes of Table 2.

Results

Table 1 presents the MSM population size estimates based on the "probability survey" method. These estimates are by geographic area for California based on the GSS point estimates of the proportion of the population that is MSM in each geographic area. Table 2 presents the MSM population size estimates based on the "never married" method. For the state as a whole, there is a substantial similarity between methods, with the survey methodology producing area estimates that are modestly larger than those produced by the "never married" method. In rural counties, the "never married" method produced a larger estimate of the MSM population size than that generated by the "probability survey" method.

Using the two estimating methods, we show in Table 3 the proportion of AIDS cases (cumulative, new diagnoses/past three years) by geographic area in relation to the proportion of MSM. Both estimating methods produce results that are proportionally parallel to the proportion of incident AIDS cases in larger MSAs, but are somewhat divergent for "other urban counties." Nevertheless, the Spearman Rank correlations between methods is over .90 and both methods

correlate over .80 with the proportionate distribution of either cumulative or incident AIDS cases.

Discussion

In general, the findings suggest that both methods provide similar estimates of the overall size of the MSM population. Further, comparing the results of these estimating methods with the geographic distribution of MSM AIDS cases suggests that both methodologies might be used to estimate the proportionate distribution of MSM across large geographic areas (i.e., recent AIDS cases provide an independent index of the relative distribution of MSM across geographic areas). The largest area discrepancy between methods is in estimating the size of the rural MSM population. This discrepancy may reflect a real concentration of never-married men in rural areas rather than an undercount of MSM. A high proportion of men that never marry in rural areas may simply reflect difficulty in finding women to marry. Additionally, a high proportion of never-married rural men could also reflect men who are attracted to a low population density area because they may prefer to be alone (including no spouse). Although the two methods are similar in estimating the size of the adult MSM population, the "probability survey" method has an advantage over the "never married" method because it can be used to collect other data useful for estimating the size of the adolescent MSM population. For example, to estimate the number of adolescent MSM in California, we obtained an estimate of the proportion of adult MSM who reported same gender sex before the age of 18 from the Urban Men's Health Study (probability sample of adult MSM in four cities [Catania et al. 2001]).[1] Approximately 60.3 percent of adult MSM report first having sex with a man prior to age 18. Multiplying this figure by the 7.5 percent of adult men in California who are MSM we obtained an estimate of the proportion of

teens that are MSM. Multiplying the proportion of MSM teens by the estimated number of male teens aged 13-17 in California from the 1999 CPS (1,284,783 adolescent males) produces a population size estimate for adolescent MSM in California of 58,104 (.603 x .075 x 1,284,783).

The associated confidence intervals of the survey-generated point estimates may also be useful for "adjusting" for under-reporting same-gender sex. For instance, the GSS point estimate and 1999 CPS generate an estimate of approximately 4,854,727 adult MSM in the United States (MSM point estimate = 5.1% (95% CI 4.4%-5.8%); U.S. male population age 18 and older = 95,190,729). The upper bound of the confidence interval (5.8%) of the point estimate, used as a conservative statistical adjustment for under-reporting, yields a national estimate of 5,521,062 adult MSM in the United States (1,147,815 MSM in California).

The lower bound of the confidence interval may be useful for purposes of costing MSM survey sample sizes (e.g., as a hedge against the limitations of using national level data to generalize to smaller geographic areas). Such a method was used with considerable success in the Urban Men's Health Study (Blair 1999). For example, using the national GSS data, the Urban Men's Health Study predicted an MSM household rate of 6.6 percent in less MSM-dense zip codes of Chicago, and obtained an actual rate of 5.5 percent. This suggests that national data may be used with some degree of predictive utility in estimating MSM populations for survey costing purposes within large urban centers if one is careful to select conservative estimates of the target population. Nevertheless, like all such estimates we must assume that we have under-

[1]Although subject to recall bias, respondents may more easily recall the age of their first major sexual encounter. This recall bias is more likely less than the bias introduced by asking teens directly if they are having same-gender sex.

represented MSM due to closetedness or discomfort in reporting sensitive information in interviews.

Closetedness may be less of an issue for urban MSM given historical changes supportive of gay culture in large urban centers. However, urban ethnic minority MSM, relative to White MSM, may be more connected and dependent on their respective local ethnic communities. Consequently, some ethnic minority men may be motivated to conceal their homosexuality if it would ostracize them from their respective ethnic communities. Thus, estimating the size of ethnic minority MSM populations may be problematic. One should not expect ethnic minority MSM to occur in proportion to the overall ethnic representation in any given area (Catania et al. 2001; Bell and Weinberg 1978; Murray 1992). This expectation assumes a simple biological model of sexual orientation without cultural effects on expression, and ignores the in-migration of White MSM to large urban gay communities.

One indirect test of the degree of under-representation of closeted MSM would be to examine if there is a more observable outcome that can be understood only if the closeted population is substantial. We can test this hypothesis by examining if HIV estimates in surveys of MSM are substantially different than estimates from public health records. If the nonresponse problem related to closetedness is large, then one should be able to detect differences, for example, in HIV rates obtained from public health records, relative to estimates obtained from probability samples of the same communities.

We have previously compared HIV prevalence data obtained from the Urban Men's Health Study's San Francisco subsample with estimates predicted from the San Francisco Department of Public Health's AIDS surveillance records (Catania et al. 2001). These estimates were similar

for the 1996-97 window[2] (HIV Prevalence: Record Data Estimates = 20-25 percent; UMHS = 20-22 percent) and suggest that under-representation of closeted MSM (and other sources of nonresponse) does not adversely affect HIV-prevalence estimates from probability-based surveys. Although the generalizability of these conclusions to MSM subpopulations and other cities is unknown, currently no evidence exists to support the contention that closetedness is more prevalent in other large urban centers.

There are several problematic issues for the current estimating methods. First, it is difficult to estimate MSM population sizes for smaller geographic areas outside large urban areas or for states that are more rural. One could assume that closetedness is more common in rural areas because conservative norms may suppress disclosure. However, this bias would have less impact on estimates at the national level or for larger, more urbanized states. Second, it is difficult to determine the time span for defining MSM. For present purposes, we defined MSM as those reporting sex with men since age 18. This definition allows respondents to be included who may be more reluctant to report current sexual activities accurately, but will report sex with men that occurred in the distant past. However, MSM who have not had sex with men as adults are excluded by the current definition we use. This is likely a very small portion of the overall MSM population. Potential under-estimation may be balanced by over-estimation due to including men who have only had sex with men once and, based on other criteria, might be considered heterosexual. Restricting definitions to men who have had sex with men recently would be restrictive and exclude men identifying as gay or bisexual who have not had sex with

[2] One may question the ability of the San Francisco Department of Public Health to accurately record AIDS cases among closeted MSM. For example, closeted MSM may misidentify themselves during HIV testing as non-MSM intravenous drug users or heterosexuals. The San Francisco data for 1996-1998 suggest that even if all men identifying as non-MSM intravenous drug users or heterosexuals were misclassified, they would still represent only 2.9 percent of new infections among men in that city (based on "Detuned Assays"). This level of complete misclassification is unlikely, so closeted men probably do not contribute significantly to HIV incidence or prevalence estimates.

men recently (e.g., men without sexual partners, or men who are very ill).

City level estimates may also need to be adjusted for migration of MSM to liberal gay communities. U.S. surveys document significant migration of MSM to large urban centers (Bell and Weinberg 1978; Murray 1992). In San Francisco, for instance, we estimated from the Urban Men's Health Study that net MSM migration over a three-year period was approximately nine percent (unpublished data, using a method that may underestimate in-migration). In-migration to large urban centers helps explain the larger proportion of MSM residing in large urban areas compared to rural and suburban areas (Binson et al. 1995). This also explains the large number of well-educated White MSM typically found in large urban areas. For example, in the Urban Men's Health Study, only nine percent of White MSM were long-term residents (greater than 10 years) compared to 39 percent of African Americans (data available from first author).

Estimating the size and distribution of the MSM population is challenging, given the low level of detail available from general data sources such as the CPS or U.S. Census. In the future, questions on sexual orientation should be added to the CPS and U.S. Census. Although this would not solve the problems of under-reporting, it would be a significant step toward acknowledging the reality of our diverse population. Until that point, however, probability surveys provide a reasonable method of estimating the size of the MSM population for adolescents and adults.

References

Bell, Alan P, and Martin S Weinberg. 1978. *Homosexualities: A Study of Diversity Among Men and Women*. New York, NY: Simon and Schuster.

Binson, D, S Michaels, R Stall, T Coates, J Gagnon, and J Catania. 1995. "Prevalence and Social Distribution of Men Who Have Sex With Men: United States and Its Urban Centers." *Journal of Sex Research* 32:245-254.

Blair, J. 1999. "A Probability Sample of Gay Urban Males: the Use of Two-phase Adaptive Sampling." *Journal of Sex Research* 36:39-44.

Catania, Joseph A., Dennis Osmond, Ronald D. Stall, Lance Pollack, Jay P. Paul, Sally Blower, Diane Binson, Jesse A. Canchola, Thomas Mills, Lawrence Fisher, Kyung-Hee Choi, Travis Porco, Charles Turner, Johnny Blair, Jeffrey Henne, Larry L. Bye, and Thomas Coates. 2001. "The Continuing HIV Epidemic Among Men Who Have Sex With Men." *American Journal of Public Health* 91:907-914.

Holmberg, Scott D. 1996. "The Estimated Prevalence and Incidence of HIV in 96 Large U.S. Metropolitan Areas." *American Journal of Public Health* 86:642-654.

Murray, Stephen O. 1992. "Components of Gay Community in San Francisco." Pp. 107-146 In *Gay Culture in America*, ed. Gilbert Herdt. Boston, MA: Beacon Press.

Table 1. Probability Survey Method: Estimating the Number of Adult MSM in CA (Data from GSS & CPS)

	A	B	A x B	
Population Size	**% Men Reporting a Male Sex Partner Since Age 18 (95% CI) [a]**	**# of Men over age 18 in CA [b]**	**Estimated # of Adult MSM (% Column)**	
Central cities of 2 largest MSAs [c]	13.0 (9.1, 16.9)	3,025,356	393,296	(45.3%)
Central cities of next 9 largest MSAs [d]	7.1 (5.1, 9.1)	1,364,665	96,891	(11.2%)
Suburbs of next 2 largest MSAs	5.8 (3.7, 7.9)	4,995,961	289,766	(33.4%)
Suburbs of next 9 largest MSAs	4.1 (2.6, 5.6)	1,489,365	61,064	(7.0%)
Other urban counties [e]	4.0 (3.1, 4.9)	591,250	23,650	(2.7%)
Other rural counties [f]	2.3 (0.9, 3.7)	142,933	3,287	(0.4%)
Total		11,609,530	867,954	(100%) (7.5%)[g]

[a] Combined data from 7 waves of the General Social Survey (1988-1991, 1993, 1994, and 1996).

[b] Data from 1999 Current Population Survey. Data weighted to agree with independent estimates of civilian non-institutional population of the United States after post-stratification by age, sex, race, Hispanic/non-Hispanic ancestry, and state of residence. Independent estimates calculated based on information from the 1990 Census; adjustment for under-coverage in the 1990 census; statistics on births, deaths, immigration, and emigration; and statistics on the size of the armed forces. The "central city" portion of an MSA is the population within the city limits of all cities in the MSA with a population >250,000.

[c] Los Angeles-Riverside-Orange County MSA (Los Angeles, Orange, Riverside, San Bernadino, and Ventura Counties) and San Francisco-Oakland-San Jose MSA (Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Santa Cruz, Solano, and Sonoma Counties).

[d] MSA (County) combinations: Bakersfield (Kern), Fresno (Fresno, Madera), Modesto (Stanislaus), Sacramento-Yolo (El Dorado, Placer, Sacramento, Yolo), Salinas (Monterey), San Diego (San Diego), Santa Barbara (Santa Barbara), Stockton-Lodi (San Joaquin), Visalia-Tulare-Porterville (Tulare).

[e] Any MSA not already mentioned, and Chico-Paradise (Butte), Merced (Merced), Redding (Shasta), San Luis Obispo-Atascadero-Paso Robles (San Luis Obispo), Yuba City (Sutter, Yuba).

[f] All counties not in MSA.

[g] Total estimated number of Adult MSM / Total population (N) of men age 18+ in California = 867,954 / 11,609,530 = 0.07476.

Table 2. Never Married Method: Estimating # of Adult MSM in CA using Holmberg's Formula (Data from CPS)

Population Size	% Men Age 18+ who are African-American (Pr_{AA})[a]	% Men Age 45+ who Never Married ($Pr_{nm > 44}$)[a]	Adjusted Percentage (Pr_{adj})[b]	Population (N) of Men Age 18+ in CA[a]	Estimated # of Adult MSM[c] (% Column)	
Central city of 2 largest MSAs[d]	11.9	12.4	11.7	3,025,356	352,823	(41.9%)
Central city of next 9 largest MSAs[d]	8.3	5.5	5.3	1,364,665	71,942	(8.6%)
Suburbs of 2 largest MSAs	4.6	5.6	5.5	4,995,961	273,339	(32.5%)
Suburbs of next 9 largest MSAs	1.1	6.8	6.8	1,489,365	100,720	(12.0%)
Other urban counties[d]	2.8	5.1	5.0	591,250	29,732	(3.5%)
Other rural counties[d]	2.1	9.1	9.0	142,933	12,870	(1.5%)
Total				11,609,530	841,426	(7.2%)[e]

[a]Data from 1999 Current Population Survey. Data weighted to agree with independent estimates of civilian non-institutional population of the U.S. after post-stratification by age, sex, race, Hispanic/non-Hispanic ancestry, and state of residence. See Table 2, footnote *a* for additional detail.

[b]$Pr_{adj} = (1 - {}^1/_2 Pr_{AA})(Pr_{nm>44})$ from Holmberg (1996).

[c](Pr_{adj} x N).

[d] MSA (County) combinations: Bakersfield (Kern), Fresno (Fresno, Madera), Modesto (Stanislaus), Sacramento-Yolo (El Dorado, Placer, Sacramento, Yolo), Salinas (Monterey), San Diego (San Diego), Santa Barbara (Santa Barbara), Stockton-Lodi (San Joaquin), Visalia-Tulare-Porterville (Tulare).

[e]Total estimated number of Adult MSM / Total population (N) of men age 18+ in California = 841,426/11,609,531 = 0.07247.

[f]Computations were made prior to rounding to avoid accumulated rounding error.

Table 3. Comparing Estimates of MSM in CA to Cumulative and Incident Cases of AIDS in CA MSM

Population Size	Survey Method (GSS-Based Estimate) %	Never Married Method (Holmberg Formula Estimate) %	Cumulative Cases of AIDS in CA MSM / MSM-IDU [a]	Incident Cases of AIDS in CA MSM / MSM-IDU [b]
2 largest MSAs	78.7 [c]	74.4 [d]	82.9%	78.0%
Next 9 largest MSAs	18.2 [e]	20.4 [f]	15.6%	20.2%
Other urban counties	2.7	3.6	0.7%	0.8%
Other rural counties	0.4	1.3	0.8%	1.0%

[a] Cumulative MSM/MSMIDU cases (n = 81,662/10,419) (California AIDS Statistics 2000) reported as diagnosed through 3/31/2000.

[b] MSM/MSMIDU cases (n = 7,571/888) (CDC AIDS Public Information Data Set 1999) reported as diagnosed in the last three calendar years for which data for the total year are available (1997-1999).

[c] Central cities and suburbs of largest MSAs from Table 1: 393,296 (45.3%) + 289,766 (33.4%) = 683,062 (78.7%).

[d] Central cities and suburbs of largest MSAs from Table 2: 352,823 (41.9%) + 273,339 (32.5%) = 626,162 (74.4%).

[e] Central cities and suburbs of next 9 largest MSAs From Table 1: 96,891 (11.2%) + 61,064 (7.0%) = 157,955 (18.2%)

[f] Central cities and suburbs of next 9 largest MSAs from Table 2: 71,942 (8.6%) + 100,720 (12.0%) = 20.6% (rounding error).